Perceptual averaging on relevant and irrelevant featural dimensions

Philip T. Quinlana, Dale J. Cohenb

and

Keith Allenc

a – Department of Psychology, University of York, UK

b – Department of Psychology, University of North Carolina at Wilmington

c – Department of Philosophy, University of York, UK

Author Note

Correspondence regarding the work should be addressed to: Philip Quinlan, Department of Psychology, The University of York, Heslington, York YO10 5DD. E-mail: philip.quinlan@york.ac.uk.

# Abstract

Here we report four experiments that explore the nature of perceptual averaging. We examine the evidence that participants recover and store a representation of the mean value of a set of perceptual features that are distributed across the optic array. The extant evidence shows that participants are particularly accurate in estimating the relevant mean value, but we ask whether this might be due to processes that reflect assessing featural similarity rather than computing an average. We set out and test detailed predictions that can be used to adjudicate between these averaging and similarity hypotheses. In each experiment, a memory display of randomly positioned bars was briefly presented followed immediately by a probe bar. Participants had to report in a Yes/No task whether the probed feature value was present. In initial experiments, we examine reports of orientation of white bars and of the color of vertical bars, respectively. Then in companion experiments we examine reports of orientation of bars whose color vary, and of the color of bars whose orientation vary. In this way, we test ideas about whether perceptual averaging occurs on a featural dimension that is irrelevant to the task. Currently, it is not known whether perceptual averaging only takes place on a task relevant dimension or whether it operates more widely.

Keywords.

Perceptual averaging, ensemble coding, ensemble perception.

Ensemble coding refers to the human visual system’s apparent ability to rapidly recover global statistical information about the content of the immediate optic array. For example, if randomly positioned texture elements are distributed throughout the visual field, an observer is able to recover a reasonably accurate estimate of the average orientation of those elements even if the presentation of the array is brief (i.e., no more than 500 ms - see for instance, Dakin & Watt, 2001; Parkes, Lund, Angelucci et al., 2001). Although there is mounting evidence that such perceptual averaging can take place on a variety of featural dimensions (see the recent review by Whitney & Yamanashi Leib, 2018), some crucial issues remain. In Experiments 1 and 2, we pit two competing explanations for the data against one another: 1) the traditional explanation that the data reflect a process of recovering and operating upon a representation that codes the never presented mean featural value (Ariely, 2001; Jeong & Chong, 2020; Watamaniuk & Duchon, 1992), or 2) the alternative explanation that the data reflect processes that are sensitive to the similarity of a probe to the *actually presented* feature values (cf. Myczek & Simons, 2008). In Experiments 3 and 4, we assess the degree to which perceptual averaging reflects attentional processing (see Chen, Zhuang, Wang, Ren & Abrams, 2021, for a recent example) by testing whether perceptual averaging takes place on a dimension that accompanies the judged dimension, but is irrelevant to the task.

Much of the extant evidence for perceptual averaging comes from studies in which participants are asked to estimate the mean value of presented features on a particular perceptual dimension such as size of a circle or orientation of a line (for instance, Khayat & Hochstein, 2018). Whereas this procedure can show the degree to which participants can extract the mean value, it provides little information about the underlying processes that make that extraction possible. A more direct method of looking at these processes is provided by the Yes/No task (Ariely, 2001; Rajendran, Maule, Franklin & Webster, 2020). Consider a key experiment by Ariely (2001) in which a ‘memory’ display, containing several circles that differed in size, was presented for 500ms. Immediately following, participants were asked to judge whether a single probe circle had been present in the memory display (i.e., respond Yes or No). The data from two participants were plotted as a function of the number of circle sizes present in the memory display and the number of circles in the display. No statistical tests are reported, but visual inspection of the respective curves indicates that the participants’ tendency to accept a novel item as present varied as a function of its distance from the mean size *regardless* of whether it had actually been present. In a separate experiment, Ariely showed that participants ‘encode quite precise information about the mean’ of sizes of the set of circles. (p. 201). Ariely (2001) concluded that the overriding propensity is to recover global attributes of a given feature set and discard information about individual items in the set: mean feature value is recovered and stored, but not the actual presented feature values.

Supporting evidence for this view has been reported in a particularly relevant follow-up study by Khayat and Hochstein (2018). In their experiments and, on each trial, participants viewed RSVP displays in which a sequence of 12 centrally presented items unfolded over time. Sequences comprised circles of various sizes, bars of various orientations or circles that varied in gray-level. Each item was presented for 100 ms separated by a blank interval of 100ms. At the end of the sequence the participant was presented with two probe items and was instructed choose which had been present in the display. Positive probes were either an old item that possessed the mean featural value of the items that had been presented or another old item. An item that possessed the mean value was known as an *Amean* item, an item that possessed a feature value within the range of those presented was known as an *A probe*, and a *B probe* was one that possessed a value outside the range of those presented. Whereas B probes were never presented in the displays, Amean and A probes may either have been presented or not. A general finding was that memory for old Amean probes was better than for old A probes and, more interestingly, accuracy of report scaled inversely with a probe’s distance from the mean (cf. Ariely, 2001). In addition, for new probe trials, participants were more accurate in rejecting the probe the further the probe was from the mean.

Amongst other things, such data as these were taken to support the view that participants automatically recovered the mean featural value of the presented items – a view that accords well with Ariely’s (2001). However, in neither case is the evidence definitive because there are two competing hypotheses that can predict that the mean value of a display will be identified more often as present than the actual presented items themselves. We call these respectively, the *Perceptual Averaging Hypothesis* and the *Similarity Hypothesis*. The extant data is generally assumed to fit most comfortably with the former hypothesis, but the latter hypothesis has never been ruled out. That is, the tendency to report the mean may arise not because of ensemble encoding that eventuates in a representation of the mean, but because the Yes/No task reflects processes that are sensitive to the similarity of the probe to the old items (where similarity is defined as distance to the nearest displayed item). By definition, because the mean probe is more similar on average to a random memory item, than is any other probe, it may be identified as present even more so than any old item.

Here, we replicate and extend the findings of Ariely (2001) and pit the two competing hypotheses against one another. Our memory display contained a randomly positioned array of 64 oriented bars (with a random half the bars of one orientation and the remaining half of another). Participants judged whether the designated feature of a probe bar was present in the display. In the first experiment, all the bars were presented in white (half of the bars were of one orientation and half were of another) and the participant had to judge whether the orientation of a probed bar was present in the memory display. The probe was either the mean of these two orientations (i.e., a mean or M probe), one of these two orientations (i.e., an old or O probe), or a new orientation that falls outside the range of the two orientations in the memory display (i.e., a novel or N probe). Assume that the bars on a trial are sampled from a series of seven possible items (numbered 1 – 7). Two such items will be selected for inclusion in the memory display. Also assume, for expository convenience, that half of the sampled bars are ‘1s’ then the remaining bars will be ‘3s’. The M probe will be a ‘2’ but the N probe can either a ‘4’, ‘5’, ‘6’, or ‘7’. Once the displayed bars are chosen then the M probe is fixed but the N probe is randomly selected from the remaining series items. By systematically varying the distance of the N probes, we aim to adjudicate between the two hypotheses.

According to the Perceptual Averaging Hypothesis, the probability of responding ‘No’ should increase directly with its distance from the mean. In addition, the mean item should be the item most erroneously categorized as being present. The Similarity Hypothesis often makes the same prediction. However, in contrast to the Perceptual Averaging Hypothesis, in our experiment the Similarity Hypothesis makes a very precise point prediction about the probability of falsely reporting the M probe as present, *p*(Yes|*M*) (see Figure 1). For this analysis, some of the N probes are critical because they are the *same* distance from the old items as the M probe. We will call these N probes, N1A and N1B. The Similarity Hypothesis assumes that these probes are as confusable with the old items as is the M probe. As such, the probability of reporting these probes as present, *p*(Yes|N1A) and *p*(Yes|N1B), provide a measure of the confusability of the M probe with the old items. If the Similarity Hypothesis is correct, the probability of confusing the M probe with the old items, *p*(Yes|*M*), should equal the joint probability of confusing either the N1A probe and/or the N1B probe with the old items (see Figure 1):

*p*(Yes|*M*)similarity = p(Yes|N1A) + p(Yes|N1B) – [p(Yes|N1A) \* p(Yes|N1B)] (1)

or, equivalently,

*p*(No|*M*)similarity = p(No|N1A) \* p(No|N1B) (2)



*Figure 1*. How predictions are derived from the Similarity Hypothesis.

Note. The first row illustrates the orientation probes. Below them is an illustration of where confusions may occur between the new probes and the old probes. Below that illustrates how the predictions arise from the data. For expository convenience, N probes here as shown as being 15o away from the nearest old item. In the experiment proper the distance between a N probe and the nearest old item varied according to which items had been selected as the old items and what the remaining items were in the series. Given these constraints, on a given trial the N probe was selected at random from the remaining items in the series (see text for further details.)

Figure 2 illustrates three hypothetical patterns of results, where *p*(Yes|*M*)similarity < p(Yes| O), *p*(Yes|*M*)similarity = p(Yes| O), and *p*(Yes|*M*)similarity > p(Yes| O). Because the Similarity Hypothesis can potentially predict all three of these outcomes (depending on the relative detectability of the old probes), it cannot be assumed that Perceptual Averaging is driving the data when *p*(Yes|*M*)similarity > p(Yes| O). Rather, it has to be established whether or not the data can be explained by the Similarity Hypothesis. Specifically, if *p*(Yes|*M*)similarity = *p*(Yes|*M*), then the Similarity Hypothesis cannot be ruled out. However, if *p*(Yes|*M*)similarity < p(Yes|*M*), then there is strong evidence for the Perceptual Averaging Hypothesis.



*Figure 2*. Three hypothetical patterns of data predicted by the Similarity Hypothesis. In all graphs p(Yes|M)similarity is presented in red, p(Yes|O) is presented in grey, and p(Yes|N1) is presented in black. p(Yes|O) remains constant in all graphs. We varied p(Yes|N1) and calculated p(Yes|M)similarity. The top graph illustrates a pattern whereby p(Yes|M)similarity < p(Yes|O), the middle graph illustrates a pattern whereby p(Yes|M)similarity = p(Yes|O), and the bottom graph illustrates a pattern whereby p(Yes|M)similarity > p(Yes|O).

Evidence from a companion experiment (Experiment 2) will be used again to try to adjudicate between these two hypotheses but this time in relation to color. In Experiment 2 all the bars in the memory display were vertically oriented, but they were colored and varied in lightness. In this regard, the experiment is a partial replication of that reported by Rajendran et al. (2020), but, in this case, it acted as a means to contrast the Perceptual Averaging and Similarity Hypotheses.

Finally, Experiments 3 and 4 address the degree to which ensemble encoding depends on attentional control. As Chen, Zhuang, Wang, Ren and Abrams (2021) have argued, the evidence is somewhat mixed on this question (see for instance, the interchange between Myczek & Simons, 2008, and, Chong, Joo, Emmanoul, &Treisman, 2008). Currently, the bulk of the evidence is restricted to cases where participants were instructed to make judgments about “attended” features on one dimension (such as the orientation of line segments, see e.g., Chen et al. 2021), and then effects of attention are gauged via the degree to which performance is affected by other “unattended” features on the same dimension. Here we examine the question of whether ensemble coding takes place on an otherwise irrelevant featural dimension. If ensemble encoding takes place automatically on all features in an array, then information from the irrelevant dimension should influence judgments on the relevant dimension and vice versa. This will be reflected in whether or not participants are seduced into committing erroneous ‘Yes’ responses if the value on the irrelevant dimension is the mean. In Experiment 3, participants judged orientation when the color of the memory items also varied across trials and, in Experiment 4, participants judged color when orientation of the memory items also varied across trials.

# Methods

## Design, stimuli and equipment

The experimental task was based on the Yes/No task described by Ariely (2001). On a given trial, a memory display was presented briefly (for 500 ms) followed by a probe display. The memory display comprised 64 randomly positioned bars and the probe display contained a single centrally presented bar. The participant was instructed to judge whether the critical feature of the probe bar had been present in the memory display and respond accordingly, either ‘Yes’ or ‘No’.

The experiment ran in a web browser via a iiyama Vision Master 505 21” color monitor. The screen resolution was set at 1600 x 1200 and the refresh rate was set at 60Hz. The bars, each 40 pixels in length 6 pixels wide, were displayed in an 800 (high) x 1000 (wide) pixel centrally positioned region. None of the memory bars were displayed within a circular region of the center (with a radius equal to the length of a bar) so as to avoid any superimposition of a memory bar with the subsequently presented central probe bar. The background color of the screen was set to gray (#808080, xyY - [0.303, 0.328, 20.234], L\*ab – [52.101, -2.774, -1.286]). The screen was gamma corrected via the use of the DataColor SpyderX Elite package and the colorimetry was undertaken with these tools.

In Experiment 1 (the Judge Orientation/Constant Color experiment), all of the bars were presented in white, but half of the memory bars were of one orientation and half were of another. The orientations to be presented were sampled from 0o to 90o separated by 15o steps. On a random half of the trials, the bars were defined relative to a rightward tilt and on the remaining half of the trials the bars were defined relative to a left-ward tilt. Prior to a given trial, two orientations from the sample were selected at random such that the chosen orientations were 30o apart. On half the trials (on 240 trials) the probe bar was one of these old items (an O probe) and the remaining trials were evenly divided between M probes (i.e., the mean of the old orientations) and N probes (i.e., the orientation was more extreme than the old orientations). On N trials the probe was selected at random from the remaining possible orientations excluding the old items and the mean. On each trial, the participant had to decide whether the probe orientation matched either of the old orientations.

In Experiment 2 (the Judge Color/Constant orientation experiment), all of the bars were vertical. However, now the lightness of the bars varied in a manner consistent with the orientation manipulation in Experiment 1. Two series of seven colors were used, one green, one blue. Across each series chromaticity was maintained (x and y were kept constant) but lightness (Y) varied (see chromaticity diagrams in the Appendix). Table 1 provides details of the two series of colors.

Table 1

*The green and blue color series and their corresponding parameter specifications*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | y | Y | L | a | b | R | G | B | Hex |
| Green Series |
| 0.322 | 0.498 | 10.759 | 39.172 | -28.659 | 29.259 | 62 | 105 | 40 | #3E6928 |
| 0.324 | 0.496 | 16.090 | 47.092 | -31.958 | 33.359 | 77 | 128 | 51 | #4D8033 |
| 0.326 | 0.492 | 23.511 | 55.595 | -34.968 | 37.306 | 95 | 155 | 63 | #5F9B3F |
| 0.325 | 0.488 | 32.637 | 63.866 | -36.799 | 38.568 | 113 | 184 | 76 | #71B84C |
| 0.323 | 0.491 | 47.588 | 74.564 | -45.063 | 46.344 | 132 | 213 | 90 | #84D55A |
| 0.320 | 0.493 | 60.685 | 82.209 | -50.536 | 50.183 | 146 | 236 | 100 | #92EC64 |
| 0.320 | 0.492 | 70.950 | 87.461 | -52.972 | 52.558 | 158 | 254 | 108 | #9EFE6C |
| Blue Series |
| 0.220 | 0.188 | 1.416 | 12.064 | 8.684 | -20.551 | 25 | 30 | 61 | #191E3D |
| 0.217 | 0.189 | 6.141 | 29.765 | 12.824 | -33.442 | 58 | 67 | 124 | #3A437C |
| 0.219 | 0.194 | 10.006 | 37.853 | 13.698 | -37.276 | 74 | 85 | 155 | #4A559B |
| 0.221 | 0.198 | 14.833 | 45.405 | 14.564 | -41.460 | 90 | 103 | 185 | #5A67B9 |
| 0.219 | 0.195 | 20.207 | 52.070 | 16.783 | -47.263 | 105 | 120 | 214 | #6978D6 |
| 0.218 | 0.193 | 25.237 | 57.306 | 18.712 | -51.811 | 118 | 135 | 239 | #7687EF |
| 0.218 | 0.193 | 28.784 | 60.591 | 19.551 | -54.133 | 126 | 143 | 253 | #7E8FFD |

*Note*. x, y, and Y measures reflect xyY parameters. L defined in CIE-L\*ab space. The RGB values are sRGB 0-255 values and HEX shows the corresponding HTML values. The colorimetry was undertaken via the DataColor SpyderX Elite package.

Half the memory bars were chosen to be of one lightness and half of another with the two lightness values separated by one intervening value (that defined the mean). On half the trials the bars were green and on half they were blue. On each trial the participant had to decide whether the probe color matched either of the old colors.

In Experiment 3 (the Judge Orientation/Varying Color experiment), both orientation and color of the probe varied even though participants were instructed to merely judge whether the orientation of the probe was present in the memory display. Now the trials were arranged according to factorial combinations of the manipulation of the orientation (M, N, O) and the manipulation of the color (M, N, O). On the O (orientation) probe trials there were 80 for each of the M, N and O color probes, and on the M and N (orientation) probe trials there were 40 for each of the M, N and O color probes.

In Experiment 4 (the Judge Color/Varying Orientation experiment), both orientation and color of the probe varied even though participants were instructed to merely judge whether the color of the probe was present in the memory display. Again, the trials were arranged around factorial combinations of the manipulation of the orientation (M, N, O) and the manipulation of the color (M, N, O). On the O (color) probe trials there were 80 for each of the M, N and O orientation probes, and on the M and N (color) probe trials there were 40 for each of the M, N and O orientation probes.

## Procedure

On each trial a small central white dot acted as an initial central fixation point presented for 500 ms. This was immediately followed by the presentation of the memory display for 500 ms. At the offset of the memory display, the probe display was presented until response. A Yes response was assigned to the ‘K’ key and a No response was assigned to the ‘D’ key.

Although reaction times (RTs) were automatically collected and participants were instructed to respond as quickly and as accurately as they could, RTs were not analyzed. Following Ariely (2001), sole interest is with the nature of the response on every trial and whether this signified a ‘Yes’ or ‘No’ response. It is not accuracy per se that is of critical interest but whether the tendency to respond ‘No’ varies according to type of probe.

The experimental scripts were written in Javascript and called the relevant jsPsych (de Leeuw, 2015) libraries. For each experiment, an initial block of 20 practice trials was presented (data from these trials were discarded prior to analysis) followed by 5 blocks of 96 experimental trials.

Participants were tested individually in a small darkened testing room containing a PC computer, a monitor, keyboard and mouse. Participants sat facing the monitor at a distance of approximately 60 cm. Responses were collected via keyboard keypresses.

## Participants

Participants were recruited via the Department of Psychology Participants panel at the University of York. The panel predominantly comprises students at the University of York. Participants were recruited via the SONA participant on-line booking tool and were offered a small payment or course credit (where appropriate) as recompense. Participants fitted the following inclusion criteria: aged between 19 and 40 yrs. old, have normal or corrected to normal vision and not have any color vision deficits.

To determine the sample size for each experiment, we ran a power simulation. Specifically, for each participant, we simulated the binomial distribution for each condition, assuming the response probabilities from a pilot study (in which PQ and KA acted as participants). We repeated this for groups of participants sized 2-10, in steps of one. We then fit our models to the data and noted the BIC and *r2* of each model fit. We determined the best fit model by the model with the lower BIC. We bootstrapped this simulation 500 times for each sample size. The results showed that we reached power of over .95 for the weakest test with 10 participants averaging 12 trials per condition (where condition was defined as steps from the mean on N probe trials). To ensure an adequate sample size, we ran 25 participants in the first experiment (more than double the number estimated by the power analysis). We used the data from the first experiment to estimate the sample size for Experiments 2-4.

# Data analysis

Participants were excluded if they were at chance with N probes that are maximally distant from the mean. The primary aim was to test the competing hypotheses. For the *Perceptual Averaging* tests, the probe trials were divided up according to the probe’s distance from the mean. For the *Similarity* tests the probe trials were divided up according to the probe’s distance from the nearest old item. The data were cast as the proportion of No responses for each probe type. The key tests compared how well the data are fit by the following log-logistic function when scored according to these two contrasting classificatory schemes,

$p\left(M\right)\_{similarity}=a+ \frac{(1-a)}{1+ e^{(b-cx)}}$ (3)

Here *a* and *b* and *c* are free parameters, and x is ordinal distance of the probe either from the mean (so as to test the perceptual averaging account) or the nearest old item (so as to test the similarity account). For each experiment the two data fits were compared using BIC.

We simply note here that Equation 3 is equivalent to,

$p\left(M\right)\_{similarity}=\frac{(1-a)}{1+ e^{(b-cx)}}$ (4)

We also tested a point prediction of the Similarity Hypothesis. Specifically, for each participant p(No|M) was calculated directly from the data and the predicted p(No|M)similarity was calculated using the N1 probes. These data were then entered into a paired sample t-test so as to determine the relation between p(No|M)similarity and p(No|M). If p(No|M)similarity = p(No|M) then the similarity hypothesis is strongly supported. In contrast, if p(No|M)similarity < p(No|M) then Perceptual Averaging Hypothesis is strongly supported. If p(No|M)similarity > p(No|M) then there is strong evidence against the Perceptual Averaging Hypothesis.

To examine the questions regarding variation on the irrelevant dimension, the accuracy scores were analyzed via ANOVA. Specifically, a two-way repeated measures ANOVA was used in which orientation value (mean, novel, old) and color value (mean novel, old) were entered as fixed factors and participants were entered as a random factor. If the irrelevant feature affects responding, then the false alarm rate should increase when the irrelevant dimension consists of either the M or the O feature values.

To assess whether Perceptual Averaging and/or Similarity of the irrelevant dimension is automated, we calculated the primary and secondary analyses describe in Experiments 1 and 2 on the irrelevant dimension.

Data collection ceased once 25 datasets had been collected per experiment that meet the inclusion criteria.

Raw data files will be deposited in a public github repository - [https://github.com/ccpluncw/ccpl\_data\_ec2021.git](https://www.google.com/url?q=https://github.com/ccpluncw/ccpl_data_ec2021.git&sa=D&source=hangouts&ust=1621515203086000&usg=AFQjCNHC2etJ09cCjzM_3TVN15RW3G4k3Q)

The project has been approved by the Ethics Committee of the Department of Psychology, The University of York.

# References

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*, 157-162. doi: <http://dx.doi.org/10.1111/1467-9280.00327>

Chen, Z., Zhuang, R., Wang, X., Ren, Y., & Abrams, R. A. (2021). Ensemble perception without attention depends upon attentional control settings. *Attention, Perception & Psychophysics, 83*, 1240-1250. doi: <https://doi.org/10.3758/s13414-020-02067-2>

Chong, S. C., Joo, S. J., Emmanoul, T.-A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics, 70*, 1327-1334. doi: <http://dx.doi.org/10.3758/PP.70.7.1327>

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*, 1-12. doi: <http://dx.doi.org/10.3758/s13428-014-0458-y>

Jeong, J., & Chong, S. C. (2020). Adaptation to mean and variance: Interrelationships between mean and variance representations in orientation perception. *Vision Research, 167*, 46-53. doi: <https://doi.org/10.1016/j.visres.2020.01.002>

Khayat, N., & Hochstein, S. (2018). Perceiving set mean and range: Automaticity and precision. *Journal of Vision, 18*, 23. doi: <https://doi.org/10.1167/18.9.23>

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics, 70*, 772-788. doi: <http://dx.doi.org/10.3758/PP.70.5.772>

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*, 739-744. doi: <http://dx.doi.org/10.1038/89532>

Rajendran, S., Malule, J., Franklin, A., & Webster, M. A. (2020). Ensemble coding of color and luminance contrast. *Attention, Perception & Psychophysics, 83*, 911-924. doi: <http://dx.doi.org/10.3758/s13414-020-02136-6>

Watamaniuk, S. N. J., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research, 32*, 931-941. doi: [http://dx.doi.org/10.1016/0042-6989(92)90036-I](http://dx.doi.org/10.1016/0042-6989%2892%2990036-I)

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69*, 105-129. doi: <http://dx.doi.org/10.1146/annurev-psych-010416-044232>

**Appendix**

*Figure A1*. Chromaticity diagram for the items in the green series.



Seven black dots are plotted to show the location of each item in the series in this color space. Most of the dots are superimposed.

*Figure A2*. Chromaticity diagram for the items in the green series.



*Note*. Seven black dots are plotted to show the location of each item in the series in this color space. Most of the dots are superimposed.

*Figure A3*. Magnified chromaticity diagram for the items in the green series.



*Figure A4*. Magnified chromaticity diagram for the items in the blue series.



*Note*. Only six dots are visible as two are superimposed.